New Three –Term CG-Method for Unconstrained Optimization

Ban A. Mitras **Nada F. Hassan**
dr.banah.mitras@gmail.com *nada.fajer10@gmail.com*
College of Computer Sciences and Mathematics
University of Mosul, Mosul, Iraq

Received on: 29/06/2011 **Accepted on: 14/12/2011**

ABSTRACT

In this paper, we proposed a new three-term nonlinear Conjugate Gradient (CG) method for solving unconstrained optimization problems. The new three-term method generates decent direction with an inexact line search under Wolfe conditions and the descent property of the new method is proved. Numerical results on some well-known test function with various dimensions showed that the new method is an efficient .

Keywords: Optimization, three term CG unconstrained optimization methods.

طريقة معامل ترافق جديدة ذات الثلاث حدود للأمثلية غير المقيدة

بان احمد متراس **ندى فالح حسن**

كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ استلام البحث: 2011/06/29 **تاريخ قبول البحث: 2011/12/14**

الملخص

تم في هذا البحث اقتراح طريقة تدرج مترافق ذات ثلاثة حدود لحل مسائل الأمثلية اللاخطية غير المقيدة. الطريقة الجديدة تولد اتجاه منحدر باستخدام خط بحث غير تام وشروط ولف. تم إثبات خاصية الانحدار للطريقة الجديدة. النتائج العددية أثبتت كفاءة الطريقة باستخدام دوال اختبار وأبعاد مختلفة ومقارنة النتائج بإحدى طرائق التدرج المترافق ذات الحدود الثلاثة.

الكلمات المفتاحية : الأمثلية، طرائق الأتجاه المترافق ذو الثلاث حدود للأمثلية غير المقيدة.

1. Introduction.

In this paper, we deal with conjugate gradient methods for solving the following unconstrained optimization problem:

Minimize f(x) …(1)

where $f:R^n \rightarrow R$ is smooth and its gradient $g(x)=\nabla f(x)$ is available. Conjugate gradient methods are very efficient for solving large-scale unconstrained optimization problems (1). For solving this problem, starting from initial guess $x_0 \in R^n$, a nonlinear conjugate gradient methods generates a sequence $\{x_k\}$as:

$$x_{k+1} = x_k + \alpha_k d_k \qquad k=1,2,\ldots \qquad \ldots(2)$$

where step size α_k is positive, which is computed by carrying out some line search, and the direction d_k is generated as:

$$d_k = \begin{cases} -g_k & for \quad k=1 \\ -g_k + \beta_k d_{k-1} & for \quad k \geq 2 \end{cases} \qquad \ldots(3)$$

In (3) β_k is known as conjugate gradient parameter. The search direction ,assumed to be a descent one which is play the main role in these methods. On the other hand, the step size α_k guarantees the global convergence in some cases and is crucial in efficiency. Plenty of conjugate gradient methods are known, and an excellent survey of these methods, with special attention on their convergence, is given by Hager and Zhang [5]. Different conjugate gradient algorithms correspond to different choices for the scalar β_k. The line search in the conjugate gradient algorithms often is based on standard Wolf conditions. The standard Wolfe conditions [9,10] are

$$f(x_k+\alpha_k d_k)-f(x_k)\le \delta\alpha_k g_k^T d_k, \quad \ldots(4)$$

$$g_{k+1}^T d_k \ge \sigma g_k^T d_k, \quad \ldots(5)$$

where d_k is a decent direction and $0 < \delta < \sigma < 1$. For some conjugate gradient algorithms, stronger Wolfe conditions defined by:

$$f(x_k+\alpha_k d_k)-f(x_k)\le \delta\alpha_k g_k^T d_k, \quad \ldots(6)$$

$$\left|g(x_k+\alpha_k d_k)^T d_k\right| \le -\sigma g_k^T d_k \quad \ldots(7)$$

are needed to ensure convergence and to enhance stability. It has been shown [9] that for FR scheme, the strong Wolfe conditions may not yield a direction of decent unless $\sigma \le 1/2$. In typical implementations of the Wolfe conditions, it is most efficient to choose σ close to one.

It is known that choices of β_k affect numerical performance of the method, and hence many researchers studied choices of β_k. Well-known formulas for β_k are the Hestenes-Stiefel (HS) [6], Fletcher-Reeves (FR) [4],Conjugate-Decent(CD) [3]and Dai-Yuan (DY) [2] formulas, which are respectively given by

$$\beta_k^{HS}=\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{FR}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{CD}=-\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \quad \beta_k^{DY}=\frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \quad \ldots(8)$$

Where $\|.\|$ means the Euclidean norm and, $y_{k-1}= g_k - g_{k-1}$.

Note that these formulas for β_k are equivalent each other if the objective function is a strictly convex quadratic function and α_k is the one dimensional minmizer.

2. Three-Term CG-Methods:

The first three-term nonlinear CG-method was presented by Nazareth [8], in which the search direction is determined by:

$$d_{k-1}=-y_k+\frac{y_k^T y_k}{y_k^T d_k}d_k+\frac{y_{k-1}^T y_k}{y_{k-1}^T d_{k-1}}d_{k-1} \quad \ldots(9)$$

The main property of d_k is that, for convex quadratic function, it remains conjugate even without ELS.

Zhang et al.[11] proposed the modified FR method (ZFR) which is defined by:

$$d_k=-\theta_k g_k+\beta^{FR}d_{k-1}, \quad \ldots(10)$$

where $\theta_k = d_{k-1}^T y_{k-1}/\|g_{k-1}\|^2$. Since this search direction satisfies $g_k^T d_k \le -\|g_k\|^2$ for all k, It can be written by the three-term form:

$$d_k=-g_k+\beta^{FR}d_{k-1}-\theta_k^{(1)}g_k, \quad \ldots(11)$$

where $\theta_k^{(1)} = g_k^T d_{k-1}/\|g_{k-1}\|^2$.

They also proposed the modified PR method(ZPR) [12] and the modified HS method(ZHS) [13],which are respectively given by:

$$d_k=-g_k+\beta^{PR}d_{k-1}-\theta_k^{(2)}y_{k-1}, \quad \ldots(12)$$

$$d_k=-g_k+\beta^{HS}d_{k-1}-\theta_k^{(3)}y_{k-1}, \quad \ldots(13)$$

where $\theta_k^{(2)} = g_k^T d_{k-1}/\|g_{k-1}\|^2$ and $\theta_k^{(3)} = g_k^T d_{k-1}/d_{k-1}^T y_{k-1}$

These three-term conjugate gradient methods which always satisfy the sufficient descent condition:

$g_k^T d_k \le -c\|g_k\|^2$, for all k,

for a positive constant c, Independently of line searches.

3. New Three-term CG-Method.

Mitras and Hassan [7] proposed a seven –parameter family defined by:

$$\beta_k^{seven} = \frac{(1-\lambda_k-\delta_k-\gamma_k-\varphi_k-\psi_k)g_k^T(y_{k-1}-v_{k-1})+\lambda_k\|g_k\|^2+\delta_k g_k^T y_{k-1}+\gamma_k\|y_{k-1}\|^2-\varphi_k^T g_k d_k+\psi_k p^T g_k}{\mu_k\|g_{k-1}\|^2+\omega_k d_{k-1}^T y_{k-1}-(1-\mu_k-\omega_k)d_{k-1}^T g_{k-1}} \quad \ldots(14)$$

where $p=(y_{k-1}-2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}})$,λ_k, $\mu_k \in [0,1]$,$\omega_k \in [0,1-\mu_k]$, $\delta_k \in [0,1-\lambda_k]$ and $\gamma_k \in [0,1-\delta k]$, $\varphi \in [0,1-\gamma_{k,}]$, $\psi \in [0,1-\varphi_k]$, parameters.(i.e. λ_k , δ_k , γ_k , φ_k and ψ_k are impossible to be equal to one at the same time; the same thing is also correct for μ_k and ω_k).

The seven-parameter family contains already existing twelve well-known formulas for β_k, so there were 27=128 cases,12 cases were succeeded,116 cases were failed.

In the present work, we derived a new three term conjugate gradient method from this family. We choose one of the failed cases that is when ($\lambda,\delta,\gamma,\varphi,\psi,\mu,\omega$) are equal to (0,1,0,0,1,0,1) respectively ,so β^{seven} yield to:

$$\beta_k = \frac{-g_k^T(y_{k-1}-\alpha_{k-1}d_{k-1})+g_k^T y_{k-1}+p^T g_k}{d_{k-1}^T y_{k-1}} \quad , \quad \ldots(15a)$$

where $p=(y_{k-1}-2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}})$ using (ILS) eq.(15a) became:

$$\beta_k = \frac{\alpha_{k-1}d_{k-1}^T g_k}{d_{k-1}^T y_{k-1}} + \frac{1}{d_{k-1}^T y_{k-1}}\left(y_{k-1}-2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}}\right)^T g_k \quad \ldots(15b)$$

Now we know that :

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$d_{k+1} = -g_{k+1} + \frac{\alpha_k d_k d_k^T}{d_k^T y_k} g_{k+1} + \frac{d_k y_k^T}{d_k^T y_k} g_{k+1} - \frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2} g_{k+1}$$

$$d_{k+1} = -[I - \frac{\alpha_k d_k d_k^T}{d_k^T y_k} - \frac{d_k y_k^T}{d_k^T y_k} + \frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2}]g_{k+1}$$

$$d_{k+1} = -Q_{k+1}g_{k+1} \text{ where } Q_{k+1} = I - \frac{d_k y_k^T}{d_k^T y_k} - \frac{\alpha_k d_k d_k^T}{d_k^T y_k} + \frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2}$$

We see Q_{k+1} is not symmetric, to symmetrize Q_{k+1} we add and subtract the term $\frac{y_k d_k^T}{d_k^T y_k}$ to get:

$$Q_{k+1}^{\#} = I - \frac{d_k y_k^T}{d_k^T y_k} - \frac{y_k d_k^T}{d_k^T y_k} + \frac{y_k d_k^T}{d_k^T y_k} - \frac{\alpha_k d_k d_k^T}{d_k^T y_k} + \frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2}$$

Use the Lipschitz condition for the third term to the numerator and denominator to get:

$$Q_{k+1}^{\#} = I - \frac{d_k y_k^T + y_k d_k^T}{d_k^T y_k} + \frac{d_k d_k^T}{d_k^T d_k} - \frac{\alpha_k d_k d_k^T}{d_k^T y_k} + \frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2}$$

$$Q^{\#}_{k+1}= I-\frac{d_k y_k^T+y_k d_k^T}{d_k^T y_k}+(\frac{1}{d_k^T y_k}-\frac{\alpha_k}{d_k^T y_k}+\frac{2\|y_k\|^2}{(d_k^T y_k)^2})d_k d_k^T$$

Hence, $Q^{\#}_{k+1}$ is symmetric, but not satisfy QN condition i.e. $Q^{\#}_{k+1} y_k \neq v_k$, to forces $Q^{\#}_{k+1}$ to satisfy QN condition we can write it as:

$$Q^{\#}_{k+1} = I-\frac{d_k y_k^T+y_k d_k^T}{d_k^T y_k}+\frac{d_k d_k^T}{d_k^T d_k}-\frac{d_k d_k^T}{d_k^T y_k}+\frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2}+(2\alpha_k-\frac{d_k^T y_k}{d_k^T d_k}-\frac{y_k^T y_k}{d_k^T y_k})\frac{d_k d_k^T}{d_k^T y_k}$$

Therefore, a direct computation shows $Q^{\#}_{k+1} y_k = v_k$ hence $Q^{\#}_{k+1}$ is symmetric and satisfy QN condition, it remains to show that $Q^{\#}$ is positive definite or equivalently the search directions generated by:

$$d_{k+1}=-Q^{\#} g_{k+1} \qquad \ldots(16)$$

are decent directions for all k. Now, our direction can be written as:

$$d_{k+1}=-(I-\frac{d_k y_k^T+y_k d_k^T}{d_k^T y_k}+\frac{d_k d_k^T}{d_k^T d_k}-\frac{\alpha_k d_k d_k^T}{d_k^T y_k}+\frac{2\|y_k\|^2 d_k d_k^T}{(d_k^T y_k)^2} +(2\alpha_k-\frac{d_k^T y_k}{d_k^T d_k}-\frac{y_k^T y_k}{d_k^T y_k})\frac{d_k d_k^T}{d_k^T y_k}))g_{k+1}$$

We call eq. (16) a new three- term conjugate gradient method. So we can write the direction of the new three-term method as follows:

$$d_{k+1}=\begin{cases}-g_{k+1} & for \quad k=0\\ -g_{k+1}+\frac{d_k^T g_{k+1}}{d_k^T y_k}y_k+\frac{1}{d_k^T y_k}(y_k^T g_{k+1}-s_k^T g_{k+1}-\frac{\|y_k\|^2 d_k^T g_{k+1}}{d_k^T y_k})d_k & for\, k\geq 1\end{cases} \qquad \ldots(17)$$

Note: Abbo has proposed formula which is nearly similar to (17). For more details See [1].

3.1 New Algorithm

Step 2. if $\|g_k\|<0$,then stop.

Step 3. Compute d_k using (17).

Step 4. Find the step length α_k satisfying (6) (7) set $x_{k+1}=x_k+\alpha_k d_k$.

Step 5. Set k=k+1, go to step 2.

4. Descent Property of the New Algorithm

The search directions generated by (17) are descent for all k if the step size satisfies Wolfe conditions.

Proof: Let $d_1=-g_1$, for $k\geq 1$ assume $d_k g_k<0$, then for k=k+1 we have:

$$d_{k+1}=-g_{k+1}+\frac{d_k^T g_{k+1}}{d_k^T y_k}y_k+\frac{1}{d_k^T y_k}(y_k^T g_{k+1}-\alpha_k d_k^T g_{k+1}-\frac{\|y_k\|^2 d_k^T g_{k+1}}{d_k^T y_k})d_k$$

$$d_{k+1}=\frac{1}{(d_k^T y_k)^2}\{-(d_k^T y_k)^2 g_{k+1}+(d_k^T y_k)(d_k^T g_{k+1})y_k+d_k^T y_k[y_k^T g_{k+1}-\alpha_k d_k^T g_{k+1}-\frac{d_k^T g_{k+1} y_k^T y_k}{d_k^T y_k}]d_k\}$$

$$\therefore d_{k+1}^T g_{k+1}=\frac{1}{(d_k^T y_k)^2}\{-(d_k^T y_k)^2 g_{k+1}^T g_{k+1}+(d_k^T y_k)(d_k^T g_{k+1})(y_k^T g_{k+1}) +(d_k^T y_k)(d_k^T g_{k+1})(y_k^T g_{k+1})-\alpha_k d_k^T y_k(d_k^T g_{k+1})^2-y_k^T y_k(d_k^T g_{k+1})^2\}$$

$$\therefore d_{k+1}^T g_{k+1} = \frac{1}{(d_k^T y_k)^2}\{-(d_k^T y_k)^2 g_{k+1}^T g_{k+1} + 2(d_k^T y_k)(d_k^T g_{k+1})(y_k^T g_{k+1}) - (d_k^T y_k)(d_k^T g_{k+1})^2 - (y_k^T y_k)(d_k^T g_{k+1})^2 \quad \ldots(18)$$

Now

$(d_k^T y_k)(d_k^T g_{k+1})(y_k^T g_{k+1}) = ((d_k^T y_k) g_{k+1})^T ((d_k^T g_{k+1}) y_k)$

Let u $= (d_k^T y_k) g_{k+1}$, v $= (d_k^T g_{k+1}) y_k$ then

$$u^T v \le \frac{1}{2}[u^T u + v^T v] = \frac{1}{2}[(d_k^T y_k)^2 g_{k+1}{}^T g_{k+1} + (d_k^T g_{k+1})^2 (y_k^T y_k)]$$

substitute in (18) we get

$$\therefore d_{k+1}^T g_{k+1} \le \frac{1}{(d_k^T y_k)^2}[-(d_k^T y_k)^2 g_{k+1}^T g_{k+1} + 2(\frac{1}{2}[(d_k^T y_k)^2 g_{k+1}^T g_{k+1} + (d_k^T g_{k+1})^2 (y_k^T y_k)]$$

$$-\alpha_k (d_k^T y_k)(d_k^T g_{k+1})^2 - (d_k^T g_{k+1})^2 (y_k^T y_k)]$$

$$\therefore d_{k+1}^T g_{k+1} \le \frac{1}{(d_k^T y_k)^2}[-\alpha_k (d_k^T y_k)(d_k^T g_{k+1})^2]$$

$$\therefore d_{k+1}^T g_{k+1} \le \frac{1}{(d_k^T y_k)}[-\alpha_k (d_k^T g_{k+1})^2]$$

By Wolfe condition $d_k^T y_k > 0$,then d_{k+1} descent for all k.

5. Numerical Result

Tables (1) ,(2), (3) and (4) are comparing between new algorithm and Zhang, Zhou and Li three–term conjugate gradient methods. The comparison involves some well-known test function with different dimensions(500,1000,5000,10000). The program is written in double precision using Fortran(2000) .The comparative Performance of the algorithm is evaluated by considering both the total number of function evaluations which is normally assumed to be the most costly factor in each iteration and the total number of iterations. The actual convergence criterion was $\|g_k\| \le 10^{-6}$. All these algorithms are implemented with the standard Wolfe line search conditions with $\rho = .001$, $\sigma = 0.9$. The results indicate that the new algorithm is more efficient.

Table(1) Numerical Comparisons between the New Method and Zhang, Zhou and Li Methods (N=500)

TEST FUNCTION	New Method NOF(NOI)	ZFR NOF(NOI)	ZPR OF(NOI)	ZHS) NOF(NOI)
1-POWELL3	63(31)	fail	42(20)	48 (22)
2-WOOD	70(31)	62 (27)	67 (30)	63 (28)
3-CUBIC	44(16)	45(16)	45(16)	44 (16)
4-SHALLOW	25 (10)	25 (10)	25 (10)	25 (10)
5-SUM	117(24)	139(24)	136(25)	107(20)
6-BELL	27(11)	27 (11)	27 (11)	29 (12)
7-ROSEN	38 (16)	76 (30)	76 (30)	76 (30)
8-RECIP	18 (6)	18 (6)	18 (6)	18 (6)
9-HELICAL	71 (32)	81 (39)	68 (32)	66 (31)
10-CANTREL	287 (33)	418 (49)	345 (39)	564(59)

Table(2) Numerical Comparisons between the New Method and Zhang, Zhou and Li Methods (N=1000)

TEST FUNCTION	New Method NOF(NOI)	ZFR NOF(NOI)	ZPR OF(NOI)	ZHS) NOF(NOI)
1-POWELL3	63(31)	fail	42(20)	48 (22)
2-WOOD	70(31)	62 (27)	67 (30)	63 (28)
3-CUBIC	44(16)	45(16)	45(16)	44 (16)
4-SHALLOW	25 (10)	25 (10)	25 (10)	25 (10)
5-SUM	119(21)	105(26)	105(26)	121(24)
6-BELL	27(11)	29 (12)	27 (11)	29 (12)
7-ROSEN	38 (16)	76 (30)	76 (30)	76 (30)
8-RECIP	18 (6)	18 (6)	18 (6)	18 (6)
9-HELICAL	71 (32)	81 (39)	68 (32)	66 (31)
10-CANTREL	287 (33)	433 (50)	345 (39)	564(59)

Table(3) Numerical Comparisons between the New Method and Zhang, Zhou and Li Methods(N=5000)

TEST FUNCTION	New Method NOF(NOI)	ZFR NOF(NOI)	ZPR OF(NOI)	ZHS NOF(NOI)
1-POWELL3	63(31)	fail	42(20)	48 (22)
2-WOOD	70(31)	62 (27)	67 (30)	63 (28)
3-CUBIC	44(16)	45(16)	45(16)	44 (16)
4-SHALLOW	25 (10)	25 (10)	25 (10)	25 (10)
5-SUM	190(31)	149(33)	149(33)	170(33)
6-BELL	27(11)	29 (12)	27 (11)	29 (12)
7-ROSEN	38 (16)	76 (30)	76 (30)	76 (30)
8-RECIP	18 (6)	18 (6)	18 (6)	18 (6)
9-HELICAL	71 (32)	81 (39)	68 (32)	66 (31)
10-CANTREL	287 (33)	522 (55)	451 (45)	617(62)

Table(4) Numerical Comparisons between the New Method and Zhang, Zhou and Li Methods(N=10000)

TEST FUNCTION	New Method NOF(NOI)	ZFR NOF(NOI)	ZPR OF(NOI)	ZHS NOF(NOI)
1-POWELL3	63(31)	fail	42(20)	48 (22)
2-WOOD	70(31)	62 (27)	67 (30)	63 (28)
3-CUBIC	44(16)	45(16)	45(16)	44 (16)
4-SHALLOW	25 (10)	25 (10)	25 (10)	25 (10)
5-SUM	125(31)	220(40)	219(38)	232(55)
6-BELL	27(11)	29 (11)	27 (11)	29 (12)
7-ROSEN	38 (16)	76 (30)	76 (30)	76 (30)
8-RECIP	18 (6)	18 (6)	18 (6)	18 (6)
9-HELICAL	71 (32)	81 (39)	68 (32)	66 (31)
10-CANTREL	287 (33)	522(55)	451 (45)	617(62)

REFERENCES

[1] Abbo, K. K., "Developing of gradient algorithms for solving unconstrained non-linear problems with artificial neural networks", University of Mosul, College of computer sciences and mathematics. Phd thesis, (2010).

[2] Dai Y.H. and Yuan Y., "A nonlinear conjugate gradient method with a strong global convergence property", SIAM Journal on Optimization, 10 (1999), 177–182.

[3] Fletcher R., "Practical Methods of Optimization", 1: Unconstrained Optimization, JohnWiley & Sons, New York, (1987).

[4] Fletcher R. and Reeves C. M., “Function minimization by conjugate gradients,” The Computer Journal, 7: 149–154, (1964).

[5] Hager W. W. and Zhang H., “A survey of nonlinear conjugate gradient methods,”Pacific Journal of Optimization, 2, pp. 35–58, (2006).

[6] Hestenes M. R. and Stiefel E., “Methods of conjugate gradient for solving linear systems,” Journal of Research of the National Bureau of Standards, Sec. B 48:409–436, (1952).

[7] Mitras, B.A. and Hassen, N.F., " New Six and Seven-Parameter Family for CG-Methods ",Iraq Journal of Statistical Science (2011). (to be published).

[8] Nazareth L, “A conjugate direction algorithm without line searches,” Journal of Optimization Theory and Applications, 23: 373–387, (1977).

[9] Wolfe P. , "Convergence conditions for ascent methods", SIAM Rev., 11: 226-235, (1969) .

[10] Wolfe P., "Convergence conditions for ascent methods". II: Some corrections, SIAM Rev.,13: 185-188, (1971).

[11] Zhang L., Zhou W. and Li D.H., "Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search', Numerische Mathematik, 104:561–572, (2006),.

[12] Zhang L., Zhou W. and Li D.H., "A descent modified Polak-Ribi`ere-Polyak conjugate gradient method and its global convergence", IMA Journal of Numerical Analysis, 26 :629–640(2006).

[13] Zhang L., Zhou W. and Li, D.H. "Some descent three-term conjugate gradient methods and their global convergence", Optimization Methods and Software, 22: 697–711(2007).